

サッカー放送映像の自動ダイジェスト化に向けた検討

平野 寿宗^{*1}、田村 徹^{*2}

A Study of Automatically Making a Digest Video from a Soccer Program

Toshizo Hirano^{*1} and Tohru Tamura^{*2}

Abstract

In this study, we investigated a method of making a digest video from a recorded broadcast of a soccer program. Digest videos are made from a compilation of match highlights – from herein referred to as ‘important scenes’ – from of a given soccer program, including, but not limited to, any goals, free kicks, substitutions of players.

First of all, we investigated what scenes should be selected for the digest video, and found that a telop, is often implemented during key scenes of a match. Therefore, we developed a method of detecting the telop by using not moving edges in adjacent frames over one second interval. Then, the detection method of the shot boundary was investigated in order to smoothly correlate and order the important scenes. To determine the shot boundary, the correlation coefficient between color histograms of the frames was used. We obtained very high accuracy of both key frame and shot boundary detection.

Finally, as a result of the user feedback retrieved via interviews, we were able to confirm a high level of satisfaction amongst viewers of the digest videos created by our proposed method.

キーワード：ダイジェスト映像，サッカー放送，テロップ検出法，ショット境界検出法，映像処理

1 まえがき

近年、デジタル技術の発展に伴って膨大な映像データがネットワーク上に蓄積され、それらを簡単に閲覧できるようになった。一方で映像データの利用者は自身が必要とする映像を効率よく探し出すことが困難となっている。その対策として、映像データを整理・分類する研究が行われている^{1) 2)}。また、映像記録メディアの大容量化、低価格化が進み、家庭でテレビ番組を大量に録画し、視聴者は好みの時間帯に録画した番組を視聴することができるようになった。しかし、大量に撮りためた録画映像をなかなか視聴する時間がないといった状況が今後想定される。そこで、いったん録画した映像を短時間に要約する研究がおこなわれている^{3) 4)}。それらの研究では、ドラマ映像や映画を対象として、物理的な構成要素、トラック構造と心理的な要素を利用した要約映像の生成手法を提案している。また、テレビで多く放送されるスポーツ映像を対象とした研究には、アメリカンフットボールの得点シーンをハイライトシーンとして検出する手法⁵⁾やサッカーのコナーキックなど特定のシーンを検出する手法に関する研究がある^{6) 7)}。これらの研究では、スポーツ放送における限られた特定シーンの検出に主眼を置いており、試合の流れを考慮しつつ視聴者が興味をもつ重要なシーンを集

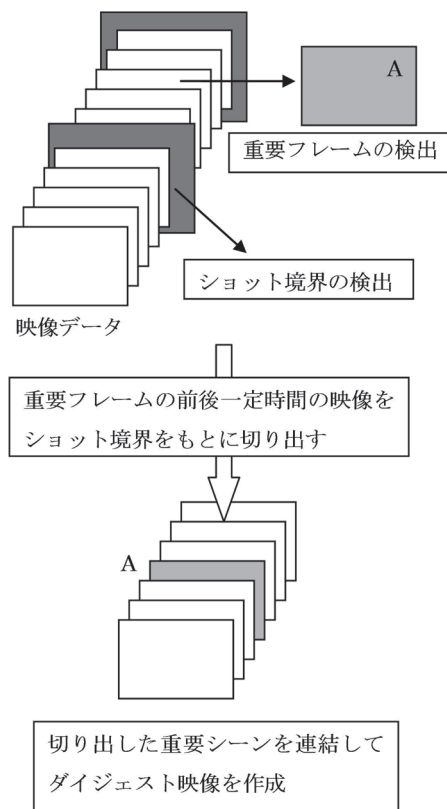


図1 ダイジェスト映像の作成手順

^{*1} 東京工芸大学メディア工学専攻 修士1年生 ^{*2} 東京工芸大学工学部メディア画像学科教授
2015年9月24日 受理

め要約することを目的とするものではない。サッカー放送の自動ダイジェスト化の研究としては、Ahmet Ekin らの研究がある⁸⁾。この研究では、ビデオソースから Shot 境界を検出し、複数のショットをからなるシーンをスローモーション、ゴール、レフリー、ペナルティーボックスに分類する。その上で、スローモーションのシーン、ゴールシーン、レフリーやペナルティーボックスのシーンを集めている。この研究もどちらかといえば、ダイジェスト映像の制作というより特定のシーンを抜き出すことに主眼が置かれている。

そこで、本研究では、サッカー放送を録画した映像から、視聴者が十分に楽しめるダイジェスト映像を作成することを目的としたダイジェスト化手法について検討することとした。サッカー放送を選んだ理由としては、サッカーは世界的に普及し、かつ人気のあるスポーツであり数多くの試合が放送されること、試合時間が決まっておき試合の最初から最後までほぼ編集なしに生放送されることから、ダイジェスト化された映像ソースの利用価値が高いと考えたためである。本研究では、ダイジェスト化する対象をサッカー放送に絞って、簡便、高精度であり、ダイジェスト化された映像を見て満足できる自動ダイジェスト化手法を提案することとした。

図1にダイジェスト映像作成方法の概要を示す。まず、映像データから重要フレームを検出する。次に重要フレームの前後一定時間をショット境界のところで切り出すことで、重要シーンとする。最後に重要シーンを繋ぎ合わせることでダイジェスト映像とする。

本論文の構成は以下のとおりである。第2章では、ダイジェスト映像としてどのようなシーンを選ぶべきか、それらに共通する画像特徴はなにかについて検討を行った結果について述べる。第3章では、重要フレームの検出法とその検出精度について述べる。第4章で、ショット境界の検出法と検出精度の確認実験の結果について述べる。第5章で、本研究の方法をもとに作成したダイジェスト映像を視聴し、その満足度を調査した結果について述べる。第6章はまとめである。

2 ダイジェスト映像に必要な場面と共通する画像

特徴

サッカー映像をダイジェスト化するにあたり、まずどのような場面が重要、つまり視聴者が見たいと感じるかを検討する。サッカーは得点を競うスポーツであることを考えると、得点に関係するシーンは最も重要と考えられる。しかし、実際にダイジェスト映像を見て楽しむためには、得点シーンだけでは不十分であることが予想される。そこで、サッカーの試合を録画した映像から、得点シーンのみを編集した映像を用意し、11名の被験者に編集した映像を視聴させた。その後、編集した映像に対して、ダイジェスト映像として見たいシーンについてコメントを収集したところ、「試合の開始、終了の場面」「選手がシュートした場

表1 重要なシーンとテロップ挿入割合

イベント	回数	テロップ	割合(%)
1) 開始・終了			
試合開始・終了、前半終了、後半開始	13	13	100
試合経過	12	12	100
2) メンバー紹介			
スターティングメンバー紹介	11	11	100
選手交代	20	20	100
ベンチ選手、監督、コーチなどの紹介	6	6	100
3) プレー			
得点	9	9	100
シュート	60	25	42
フリーキック(ゴール付近)	12	5	42
コーナーキック	22	3	14
反則(イエロー、レッドカード)	6	6	100

面」「選手の紹介」「フリーキックやコーナーキック」「選手交代」「反則」などがダイジェスト映像に必要な重要な場面であることがわかった。

そこで、上述したダイジェスト映像に必要な場面に共通する画像特徴について録画映像をもとに調べたところ、重要な場面の多くで画面下部に得点経過や選手の名前などを示すテロップがインポーズされることがわかった。実際の例を図2に示す。図2の例は上が得点シーン、下が選手交代のシーンであり、いずれも画面の下部にテロップが挿入されていることがわかる。



図2 重要なシーンの例。
上：得点シーン、下：選手交代

サッカー放送を録画した3試合分の映像から、イベントとテロップの挿入割合について調べた結果を表1に示す。回数とは映像中にそのシーンが現れた回数であり、テロップとは、そのうち図2のようにテロップが挿入されていた回数を表している。試合の開始やメンバー紹介時は100%テロップが挿入されていた。また、得点、イエローカードまたはレッドカードが提示される反則シーンも100%テロップが挿入された。シュートやゴールに近い位置でのフリー

キックは 42%、コーナーキックは 14%でテロップが挿入されていた。そこで、画面にテロップが挿入されたシーンを集めることで、視聴者が見たいと感じるシーンの多くを集めることができると考えられる。その他、対戦チームの紹介や試合日程、監督や審判の紹介などにもテロップが表示された。試合経過とは直接関係ないものの、このような情報も視聴者を満足させるといった面から必要であると考えられる。

そこで、本研究ではダイジェスト映像に必要な場面を示すフレームの画像特徴を「画面の下部にテロップが存在する」とし、この条件をみたすフレームを含むシーンを検出することとした。

3 フレーム画像中のテロップ領域の検出

本章では、ダイジェスト映像に必要なシーンを抜き出すことを目的とした、フレーム画像中のテロップの検出法について述べる。

3.1 テロップ検出法

スポーツ映像内のテロップの検出法として、ハフ変換を用いて画像中の水平線を検出する方法⁹⁾、テロップ内の数字を検出する方法¹⁰⁾などが提案されている。また、一般的な映像内におけるテロップの検出法としては、連続する2枚のフレーム画像の静止エッジの分布を利用した方法が提案され、良い検出精度を得ている¹¹⁾。

本研究では、ある時間間隔はなれた2枚のフレーム画像の静止エッジの分布をもとに、フレーム画像内のテロップを検出することにする。その際、サッカー放送をダイジェスト化するという研究目的にそって、以下の点を新たに工夫することで、処理コストの低減や検出精度を高めることとした。

1) テロップ探索範囲を画面の下半分の領域とすることで探索効率を向上する。これは、ダイジェスト映像に必要なシーンに現れるテロップはそのほとんどすべてが画面下の領域に表示されるためである。

2) 静止エッジの検出を行う2枚のフレーム時間間隔を1秒とした。これはサッカーの試合におけるテロップ表示時間は短くとも数秒は表示されるため、1秒ごとにフレーム画像を処理することで、テロップの検出が可能であるためである。これにより連続するフレーム間で検出をおこなう場合に比べ、処理コストを減らすことができる。

3) テロップ候補領域を設定した。テロップが存在する場合には、画面の小さな領域に静止エッジが集中することを利用する。具体的には後述するように、画面サイズの10%以下の領域内に画面内の全静止エッジの80%以上が含まれる場合に、テロップ候補領域とした。このようにすることで、たとえば、固定カメラでフィールド全体を映した場面やゴールポストや動きの少ない対象を映した場合など画面の比較的広い範囲に静止エッジが分布する場合に検出対象から除外することができ、テロップの誤認識を

減らすことができる。

4) テロップ内には文字が水平方向に並ぶことを利用し、文字の大きさに相当する正方領域を設定し、正方領域内の静止エッジの割合と連続性を評価し、テロップ領域を検出する。

次に、具体的なテロップの検出法について述べる。図3にテロップ検出の主な処理の流れを示す。

(1) 静止エッジの検出

ある時点でのフレーム1とフレーム1から1秒時間をおいたフレーム2から静止エッジを検出する。まず、それぞれのフレーム画像から3×3のSobelフィルタによって、エッジを検出し、エッジ画像とする。その後、両エッジ画像のANDをとることによって静止エッジ画像を得る。サッカー放送において検出するべきテロップは画面下半分に位置することが多いため、この静止エッジ画像はフレーム画像の下半分とした。

(2) テロップ候補領域の決定

テロップを検出する際の特徴点として、静止エッジを利用する場合、静止した背景物体のエッジが誤認識の原因となる可能性がある。この誤認識を減らすためにエッジの大局的な分布を利用して、テロップ候補領域を設定する。フレーム画像内に検出するべきテロップが存在する場合には、静止エッジがある狭い領域に集中する。そこで、静止エッジ画像のエッジを縦横それぞれの方向で積算し、静止エッジ総数の10%から90%の範囲を求めた。求めた領域がフレーム画像の大きさの10%より小さく、かつ求めた範囲内の全pixelの5%以上が静止エッジである場合に、テロップ候補領域とした。

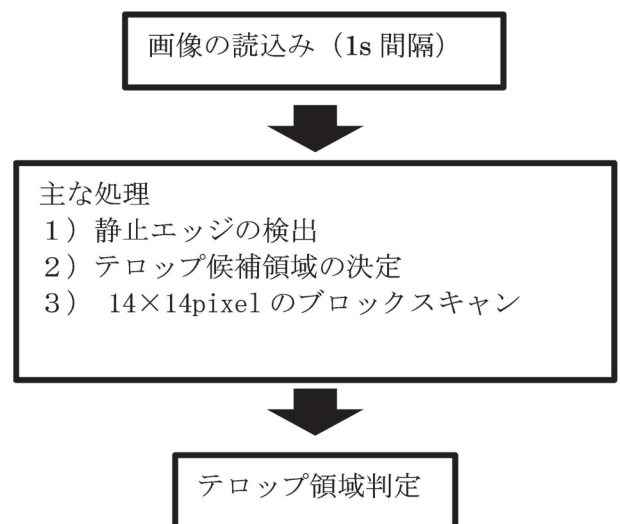


図3 テロップ検出処理

(3) ブロックスキャン

縦横14pixelの正方領域を考え、静止エッジ画像をスキャンしながら正方領域内の静止エッジの数をカウントする。ブロックスキャンは水平、垂直方向に7pixelずらし

ながら行った。正方領域内の静止エッジ数の割合が 35% より大きいときにテロップである可能性があるとした。

(4) テロップの判定

(3) のブロックスキャンの結果得られた、正方領域が水平方向に 3 領域分連続し、かつ (2) で検出したテロップ候補領域内に存在する場合にテロップ領域として検出した。

3.2 検出実験

テロップ領域の検出精度を確認するために、実際のサッカー放送を録画した映像ソースから、1 秒の間隔で 2 枚のフレーム画像を取り出し、そのフレーム画像を用いてテロップ検出実験を行った。テスト画像は放送の中で表示された様々なテロップを含む 40 対とテロップを含まない 60 対の計 100 対をテスト画像とした。検出実験は異なる 3 試合の録画映像それぞれ 100 対、計 300 対について行った。

結果を表 2 に示す。再現率と適合率は以下の式より計算した。検出実験では 3 試合の合計で再現率 0.92、適合率 0.97 と高い検出精度を得ることができた。

テロップの誤検出によって関心度の低い不要なシーンがダイジェスト映像に紛れ込むことを極力少なくするといった点を重視すると、特に高い適合率が望まれる。

$$\text{適合率} = \frac{\text{正検出数}}{\text{正検出数} + \text{誤検出数}}$$

$$\text{再現率} = \frac{\text{正検出}}{\text{テロップ入り画像数}}$$

静止エッジを用いた際に誤検出が生じる場合の多くは、カメラアクションがなく 2 枚のフレーム画像で背景部分に移動がない場面が多い。このような場面では、画面内に多くの静止エッジが存在することとなり誤検出の原因となる。本研究では、静止エッジの大局的分布領域によるテロップ候補領域を設定したことが誤検出の低減に大きな効果をあげた。図 4 にテロップ候補領域の設定によって正しくリジェクトできた例を示す。図 4 の上は、フィールド全体を俯瞰するようにカメラを固定して撮られた映像である。ブロックスキャンの結果からは数多くのテロップ領域が存在しており (図中に小さな四角として表示)、このままでは誤検出となる。しかし、同時に静止エッジは画面の広い領域に分布するため、テロップ候補領域 (図中に大きな四角として表示) が大きくなる。テロップ候補領域の大きさの判断基準により、正しくリジェクトすることができた。図 4 の下はゴールの後ろからカメラアクションがない状態で撮られた映像で、上の場合と同様にテロップ候補領域を設定することで正しくリジェクトすることができる。

4 ショット境界の検出法

映像のショット境界を検出する方法としては、2 枚の連

	正検出	誤検出	再現率	適合率
試合 1	38	2	0.95	0.95
試合 2	37	1	0.93	0.97
試合 3	35	0	0.88	1.00
計	110	3	0.92	0.97

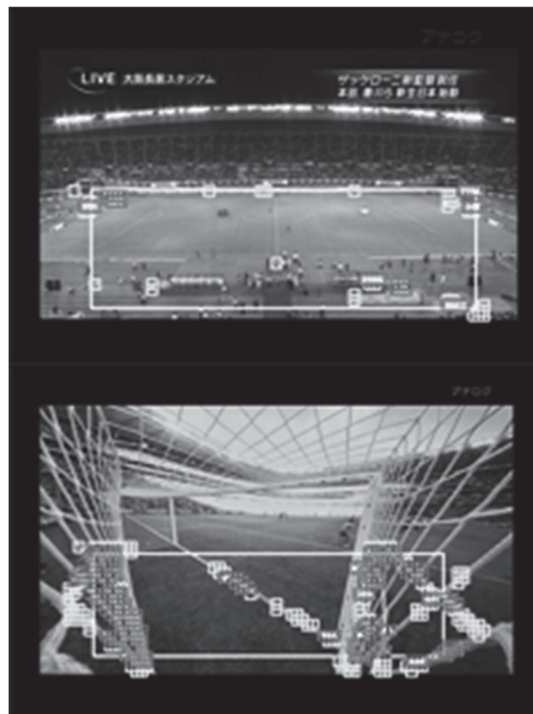


図 4 正しくリジェクトした例

続するフレーム画像間の色ヒストグラムの類似性を特徴量として用いる方法が提案されている。Ahmet は、2 枚のフレーム画像の色ヒストグラムを求め、その差をもとに 2 枚のフレーム画像の間にショット境界があるか否かを評価している。その際、フレーム内の緑色の領域の大小によってショット境界の判定に用いる色ヒストグラムの類似性の閾値を変えている。こうすることでフィールドを俯瞰するような緑の多いシーンと、たとえば、観客席など緑の少ないシーンによって異なる判定基準を用いたショット境界の検出法を提案している⁸⁾。新田らは、フレーム画面を 16 の小領域に分割し、2 枚のフレーム間で隣接するブロック間の色ヒストグラムの差分和をもとにアメリカンフットボールの試合映像のショット領域の検出方法を提案している¹²⁾。また、フレーム間の色ヒストグラムの類似性と動きベクトルを組み合わせた方法^{13) 14)} や物体追跡手法と組み合わせた手法¹⁵⁾ などが提案されている。Ahmet らの方法の場合、シーンの大まかな内容、フィールドを映しているのか、それ以外の場所を映しているのかを区別することで、ショットの検出精度の向上を図っている。しかし、この方法は、フレーム内の選手の動きやカメラのパンやズームといった動きを考慮していない。一方、新田らの手法の場合は、フレームを小領域に分割し、それぞれ隣接する領域との色ヒストグラムを比較することで、フレーム

内の選手の動きやカメラアクションへの対応が図られている。しかし、フレームを小領域に分割し、すべての小領域ごとにその周辺小領域との間で色ヒストグラムの類似度を計算する必要があり、計算コストの面で検討の余地がある。また、フレームの分割方法は固定されており、フレーム内の選手の大きさや位置と小領域が必ずしも一致するとは限らない。また、動きベクトルや物体追跡を合わせて利用することで、カメラアクションへの対応を図ることができるが、新田らの手法と同様にその分、計算コストは大きくなる。

そこで、本研究ではフレーム全体の色ヒストグラムの類似性と、フレーム内でもっとも大きな領域の色を除いた残りの色ヒストグラムの類似性の両者を利用したショット境界の検出法を提案する。たとえば、図5のような場合では、緑のフィールドが多くを占め、フィールド全体の色ヒストグラムは高い類似性を持つ。一方で、フレーム内でもっとも大きな領域、図5の場合にはもっとも大きな領域は緑色でフレーム画像の背景を構成している、最大領域の色を除いた残りの色ヒストグラムの類似度は低い値となる。図5の場合は上下のフレーム間にショット境界があり、フレーム全体の色ヒストグラムともっとも大きな領域を占める色を除いた色ヒストグラムの類似度を併用することで、ショット境界の検出を正しく行うことができる。同一色で大きな領域は、フレーム画像の背景を構成する 경우가多く、簡易的にフレーム内を背景と前景に分離することが



図5 類似した背景で別ショットの例

でき、選手がフレーム内で移動した場合とショットが切り替わって異なる選手を映した場合を区別でき、カメラのパンやズームといった動きにも対応することができる。また、フレーム全体の色ヒストグラムを一度計算するだけで、2つの類似度を簡単に求めることができ、計算コストも少なく済むといったメリットが考えられる。

4.1 2つの色ヒストグラム

ショット境界の判定は、1秒間隔離された2枚のフレーム画像の色ヒストグラムの相関係数を用いて行う。図6に1秒間隔離された2枚のフレーム画像とそのフレーム画像から計算した色ヒストグラムを示す。図6の例では、2枚のフレーム画像は異なるショットに属しており、2枚のフレーム画像の間にショット境界が存在することになる。色ヒストグラムの計算では、まず、フレーム画像の各画素値(R、G、B)から、その画素の色相値(H)に変換した。変換したHの値域を16等分して16色の色相とした。フレーム画像の各画素からHの値を計算し、色相ごとに度数を求め、最大度数を255に正規化した。

ここで、連続する2枚のフレーム間でショット境界の判定を行わないのは、例えば、ショットが数フレームの間で徐々に切り替わるような場合にも対応可能であり、すべての連続するショット間で色ヒストグラムの相関係数を計算する場合と比較すると計算コストを低く抑えることができるためである。一方このままでは、正確なショット境界をフレーム単位で決定することはできないが、ダイジェスト映像のシーンの変わり目で違和感を感じないといった、本研究の目的からすれば、大まかなショット境界が検出できれば十分である。また、必要となれば、ショット境界が検出された場合のみ、その間の連続する2枚のフレーム画像をとりだし、同様の判定を行うことで、より正確なショット境界の位置を検出することも可能であると考えている。

本研究では、色ヒストグラムとして、フレーム画像全体の色ヒストグラム(全体的色ヒストグラムと呼ぶことにする)と第1フレームの最大色相を除いた残りの色ヒストグラム(ここでは部分的色ヒストグラムと呼ぶことにする)の2つの色ヒストグラムを用いてショット境界の判定を行う。図6の例では、部分的色ヒストグラムとは、第1フレームの最大色相、つまりもっとも左端のbinを両フレームのヒストグラムから除いた15色相を用いることになる。

2枚のフレーム画像の色ヒストグラムの類似性の評価値として、フレーム1の各色相の度数とフレーム2の各色相の度数との間の相関係数を用いることにした。



図6 1秒間隔の2枚のフレーム画像(上)とその色ヒストグラムの計算結果(下)

4.2 相関係数の計算

全体的色ヒストグラムの相関係数(全体色の相関係数)と部分的色ヒストグラムの相関係数(部分色の相関係数)を用いて、ショット境界の有無を判定する。ここでは、ショット境界の有無を判定する基準を決定するために、複数のフレーム画像対から相関係数の計算を行う。

フレーム画像対として、異なる3試合のサッカー放送を録画したものから、試合ごとにショット境界を含む50組と、含まない50組、3試合で150組ずつ、計300組を用意した。フレーム画像対の時間間隔は1秒とした。図7にフレーム画像対の例を示す。上がショット境界を含む場合、下はショット境界を含まない場合の例である。



図7 フレーム画像対。ショット境界を含む(上) ショット境界を含まない(下)

図8にショット境界を含まないフレーム画像対150組とショット境界を含むフレーム画像対150組から計算した相関係数の値を示す。横軸が全体色の相関係数、縦軸が部分色の相関係数の値である。●がショット境界を含まないフレーム画像対の結果であり、○はショット境界を含んだフレーム画像対の結果である。ショット境界を含まないフレーム画像対は、図の右上、つまり全体色の相関係数、

部分色の相関係数ともに高い値を示す場合が多いことがわかる。一方、ショット境界を含んだフレーム画像対の場合には、左下、つまり全体色の相関係数、部分色の相関係数ともに低い値を示すものが多い傾向がみられる。

全体色と部分色2つ相関係数を使うことによって以下の効果を得ることができる。図5に示すように背景が似ているが、前景が異なるような場合、つまりショット境界を含むフレーム画像対が似たような構図でありながら、異なる人物をとらえたような場合が典型的例となる、図5の場合では、全体色の相関係数が0.99、部分色の相関係数が0.20であった。このような場合、全体色の相関係数だけで、ショット境界の判定正しく行うことはできないが、部分色の相関係数と合わせて判定することで、ショット境界を正しく検出できる可能性がある。

また、図9の上のように、同じショットでカメラがパンしたり、ズームしたりするよう場合には、背景が異なり、前景が似ているフレーム対となる。このような例では全体色の相関係数が低くなるが、部分色の相関係数は高い値となる傾向がある。図9の上の場合には全体色の相関係数が0.28、部分色の相関係数が0.88であった。この場合、全体色の相関係数のみによってショット境界の判定を行うと誤検出となるが、部分色の相関係数も合わせて判定に用いることで、誤検出を防ぐことができる可能性がある。

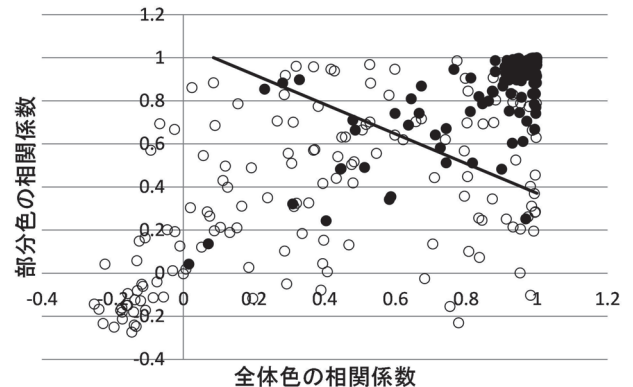


図8 相関係数の計算結果
● : ショット境界を含まない
○ : ショット境界を含む
- : 式1の直線

一方、図9の下例のようにフィールドが画面の多くを占めるような構図から、選手を含みつつ、フィールドが占める割合が少ない構図へとショットが切り替わった場合などは、全体色の相関係数が0.32と低くなる。しかし、最初のフレームの最大色、つまりフィールド部分を除いた、部分色の相関係数が0.96と高くなる。この場合には、全体色の相関係数のみによって判定すれば正しく判定できるが、部分色の相関係数を合わせて使うことで、未検出となる可能性がある。



図9 全体色の相関係数が低く、部分色の相関係数が高い例。ショット境界を含まない(上)、ショット境界を含む(下)。

4.3 ショット境界の判定基準

ここでは、図8に示した、相関係数の計算結果をもとに、ショット境界の判定基準について検討する。本研究では、図8に示した、300組のフレーム画像対の全体色の相関係数と部分色の相関係数の値を説明変数とし、ショット境界を含むか否かを目的変数とした線形判別分析を行った。分析結果を表3に示す。全体色の相関係数、部分色の相関係数ともにショット境界の有無の判定に対して有意である。

そこで、表3の判別係数をもとに、判別式を以下のように定めた。

$$y = -2.898r_1 - 4.217r_2 + 4.464 \dots (1)$$

ここで、 r_1 は全体色の相関係数、 r_2 は部分色の相関係数である。 y の値が正となる場合は、2枚のフレーム画像対の間にショット境界があると判定し、 y の値が負の場合には、ショット境界がないと判定する。判別式を図8に直線で示す。直線より上の領域はショット境界がないと判定する領域、直線より下はショット境界があると判定する領域となる。図8の300対のデータに対し、適合率と再現率を計算すると、適合率は0.90、再現率は0.76となった。

	判別係数	標準判別係数	
全体色の相関係数	-2.898	-1.165	
部分色の相関係数	-4.217	-1.630	
定数項	4.464		
	F値	p値	判定
全体色の相関係数	25.14	0.000	[**]
部分色の相関係数	49.19	0.000	[**]
定数項			

4.4 判別精度の確認実験

ここでは、式(1)を用いたショット境界の判定精度について検討する。判定精度を検討するために、サッカー放送を録画した2試合の映像から、1分間のシーンを各試合15シーン、2試合で30シーン、すなわち30分の映像を用意した。映像データの先頭フレームから順に1秒間隔で2つのフレーム画像対を抜き出し、全体色の相関係数と部分色の相関係数をもとに、式(1)を用いてショット境界の有無を判定した。使用した映像ソースは式(1)を決定する際に使用した試合とは異なる試合の映像である。

表4に結果を示す。ショット境界の数は試合1が179、試合2が174、2試合合計で353であった。そのうち式(1)によって正しく検出された数(正検出)は、試合1が133、試合2が137、ショット境界ではないところでショット境界と判定した誤検出は試合1で4、試合2で13であった。また、ショット境界の検出に失敗した未検出数は試合1で46、試合2で37であった。適合率と再現率を計算した結果を見ると、試合によって、多少の違いがあるが、全体では、適合率が0.94、再現率が0.76となった。式(1)を決定した300組のデータから得られた適合率、再現率と比較すると、再現率はほぼ同じ値となったが、適合率がやや高くなった。これは、式(1)を決定する際に、誤検出を起しそうな画像対を多めに選んだことが影響しているものと考えている。本研究の目的はダイジェスト映像の制作であり、ショット境界ではないところにショット境界を検出して重要なシーンをショットの途中でつなぎあわせてしまう不自然さを少なくすることが求められる。そこで、誤検出を少なくすることは未検出を少なくするより重要な課題と考えて、ショット境界の判別式をたてる際に誤検出を起しやすそうなフレーム画像対をあらかじめ多めにいれておくようにした。

	境界数	正検出	誤検出	未検出
試合1	179	133	4	46
試合2	174	137	13	37
計	353	270	17	83
	適合率	再現率		
試合1	0.97	0.74		
試合2	0.91	0.79		
計	0.94	0.76		

5 ダイジェスト映像の評価

ここでは、上記の手法にそって作成したダイジェスト映像を実際に視聴し、その満足度について調べた結果について述べる。ダイジェスト映像は以下の基準によって手動で作成した。

ダイジェスト映像の作成法：画面の下半分にテロップが表示されたフレームから30秒を遡りそこから45秒間の映像を抜き出し、それらをつなぎ合わせてダイジェスト映像とした。

実験素材には、放送局の異なる2試合を録画した2時間

程度の映像を使用した。ダイジェスト化された映像の時間は素材1が28分16秒、素材2が38分であった。それぞれのダイジェスト映像を素材1は9名の大学生、素材2は8名の大学生に視聴させて、ダイジェスト映像として満足度を100点満点で採点したところ、平均として素材1の満足度が78点、素材2が71点の評価を得た。

本研究が示すダイジェスト化手法によって作成されるダイジェスト映像が視聴者に対して一定の満足度を与えることが確認できたと考えている。

6 むすび

本研究では、世界的に人気があり、テレビで放送される機会が多いサッカー放送映像を自動的にダイジェスト化する手法について検討した。ダイジェスト映像としてまとめるシーンには、得点シーンだけではなく、試合の開始や終了、選手紹介、選手交代、反則、フリーキック、コーナーキックなど多様なシーンを含むとしたうえで、それらのシーンに含まれる重要フレームの検出法について検討した。重要フレームの検出では、フレーム画像の下半分にテロップが表示されることを画像特徴として、テロップの検出を行った。テロップの検出では1秒間隔離れた2枚のフレーム間の静止エッジを用い、静止エッジの大局的分布領域の大きさと文字の大きさに相当する領域のブロックスキヤンの両者を用いた判定法を提案し、良好な検出結果を得た。次に、重要シーンをショット境界に合わせて切り出すためにショット境界の検出法について検討した。ショット境界の検出には、全体色の相関係数と部分色の相関係数を用いることで、カメラアクションへの対応と計算コストの低減を同時に図った。ショット境界の判定には、線形判別分析を用いることで最適化をはかり、良好な検出結果を得ることができた。最後に、本研究の手法によって作成したダイジェスト映像を実際に視聴したところ、一定の満足感を与えることが確認できた。

以上、本研究では、テレビなどで放送される2時間程度のサッカー放送を試合開始や終了、試合中の重要なシーンを含み、30分程度にまとめたダイジェスト映像の自動作成法として有効な手法を提案した。本研究の成果は、たとえば、ホームビデオレコーダー等に搭載し、仕事や外出などで、試合のテレビ放送を視聴できなかった人が、あとで、試合の概要を短時間で楽しむ場合などに活用できると考えられる。

文献

- 1) Y. Li、田中譲、“メタデータの管理に基づくビデオデータベースの構成”、情処学論、vol.39, no.4, pp.1137-1145、(1998)
- 2) Makoto YAMAMOTO, Miki HASEYAMA, “An Accurate Scene Segmentation Method Based on Graph Analysis Using Object Matching and Audio Feature”, IEICE TRANS, FUNDAMENTALS, Vol.E-A, No8, pp.1883-1891 (2009)
- 3) 森山剛、坂内正夫、“ドラマ映像の心理的内容に基づいた要約映像の生成”、電子情報通信学会論文誌、Vol.J84-D-

- II No.6, pp.1122-1131、(2001)
- 4) 石井孝和、吉高淳夫、平川正人、市川忠男、“映画の文法に基づくビデオ画像の内容検索”、情処学研報、97-DBS-111, pp.65-72、(1997)
- 5) 宮内進吾、馬場口登、北橋忠宏、“テキスト・音声・画像の協調的処理による放送型スポーツ映像におけるハイライト検出とインデキシング”、電子情報通信学会論文誌、Vol.J85-D-II No.11, pp.1692-1700、(2002)
- 6) 丸尾二郎、岩井儀雄、谷内田正彦、越後富夫、飯作俊一、“サッカー映像からの特定イベントの抽出”、信学技報、PRMU99-41 (1999)
- 7) 三須俊枝、苗村昌秀、藤井真人、八木伸行、“選手フォーメーション解析に基づくサッカーイベント判別法”、映像情報メディア学会誌、Vol.61, No.9, pp.1367-1375、(2007)
- 8) Ahmet Ekin, A. Murat Tekalp and Rajiv Mehrotra, “Automatic Soccer Video Analysis and Summarization”, IEEE TRANS. IMAGE PROCESSING, VOL.12, NO.7, pp.796-806、(2003)
- 9) Dian Tjondronegoro, Yi-Ping Phoebe Chen and Binh Pham, “Integrating Highlights for More Complete Sports Video Summarization”, IEEE MultiMedia, pp.22-37, October-December 2004.
- 10) 田村 徹、陳 曉琴：サッカー放送映像のダイジェスト化に向けたキーフレーム検出、第10回情報科学技術フォーラム講演論文集、第3分冊、pp.447-452 (2011)
- 11) 新井啓之、桑野秀豪、倉掛正治、杉村利明、“映像中のテロップ表示フレーム検出方法”、電子情報通信学会論文誌、Vol.J83-D-II No.6, pp.1477-1486、(2000)
- 12) 新田直子、馬場口登、北橋忠宏、“放送型スポーツ映像の構造を考慮した重要シーンへの自動アノテーション付け”、電子情報通信学会論文誌、Vol.J84-D-II No.6, pp.1838-1847、(2001)
- 13) 鈴木賢一郎、中嶋正臣、坂野鋭、三部靖夫、大塚作一、“動き方向ヒストグラム特徴を用いた映像データからのカット点検出法”、電子情報通信学会論文誌、Vol.J86-D-II No.4, pp.468-478、(2003)
- 14) Bo Han, Yichuan Hu, Guijin Wang, Weiguo Wu and Takayuki Yoshigahara, “Enhanced Sports Video Shot Boundary Detection Based on Middle Level Features and a Unified model”, IEEE TRANS. on Consumer Electronic, VOL.53, NO.3, pp.1168-1176, (2007)
- 15) 田村 徹、陳 曉琴、“サッカー放送映像のダイジェスト化に向けたショット境界の検出”、2013年電子情報通信学会総合大会講演集